HUMAN REGULATORY MOLECULES

This application is a divisional of USSN 09/518,865 filed 3 March 2000, which was a divisional of USPN 6,132,973, issued 17 October 2000, which was a divisional of USPN 5,932,442, issued 3 August 1999.

FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of human regulatory molecules which are implicated in disease and to the use of these sequences in the diagnosis and treatment of diseases associated with cell proliferation.

BACKGROUND OF THE INVENTION

Cells grow and differentiate, carry out their structural or metabolic roles, participate in organismal development, and respond to their environment by altering their gene expression. Cellular functions are controlled by the timing and amount of expression attributable to thousands of individual genes. The regulation of expression is metabolically vital in that it conserves energy and prevents the synthesis and accumulation of intermediates such as RNA and incomplete or inactive proteins when the gene product is not needed.

Regulatory protein molecules function to control gene expression. These molecules turn individual or groups of genes on and off in response to various inductive mechanisms of the cell or organism; act as transcription factors by determining whether or not transcription is initiated, enhanced, or repressed; and splice transcripts as dictated in a particular cell or tissue. Although regulatory molecules interact with short stretches of DNA scattered throughout the entire genome, most gene expression is regulated near the site at which transcription starts or within the open reading frame of the gene being expressed. The regulated stretches of the DNA can be simple and interact with only a single protein, or they can require several proteins acting as part of a complex in order to regulate gene expression.

The double helix structure and repeated sequences of DNA create external features which can be recognized by the regulatory molecules. These external features are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which cause distinct bends in the helix. Such features provide recognition sites for the binding of regulatory proteins. Typically, these recognition sites are less than 20 nucleotides in length although multiple sites may be adjacent to each other, and each may exert control over a single gene. Hundreds of these DNA sequences have been identified, and each is recognized by a different protein or complex of proteins which carry out gene regulation.

The regulatory protein molecules or complexes recognize and bind to specific nucleotide sequences of upstream (5') nontranslated regions, which precede the first translated exon of the open reading frame (ORF); of intron junctions, which occur between the many exons of the OR; and of downstream (3') untranslated

regions, which follow the ORF. The regulatory molecule surface features are extensively complementary to the surface features of the double helix. Even though each individual contact between the protein(s) and helix may be relatively weak (hydrogen bonds, ionic bonds, and/or hydrophobic interactions) and the 20 or more contacts occurring between the protein and DNA result in a highly specific and very strong interaction.

Families of regulatory molecules

Many of the regulatory molecules incorporate one of a set of DNA-binding structural motifs, each of which contains either α helices or ß sheets and binds to the major groove of DNA. Seven of the structural motifs common to regulatory molecules are helix-turn-helix, homeodomains, zinc finger, steroid receptor, ß sheets, leucine zipper, and helix-loop-helix.

The helix-turn-helix motif is constructed from two α helices connected by a short chain of amino acids, which constitutes the "turn". The two helices interact with each other to form a fixed angle. The more carboxy-terminal helix is called the recognition helix because it fits into the major groove of the DNA. The amino acid side chains of the helix recognize the specific DNA sequence to which the protein binds. The remaining structure varies a great deal among the regulatory proteins incorporating this motif. The helix-turn-helix configuration is not stable without the rest of the protein and will not bind to DNA without other peptide regions providing stability. Other peptide regions also interact with the DNA, increasing the number of unique sequences a helix-turn-helix can recognize.

Many sequence-specific DNA binding proteins actually bind as symmetric dimers to DNA sequences that are composed of two very similar half-sites, also arranged symmetrically. This configuration allows each protein monomer to interact in the same way with the DNA recognition site and doubles the number of contacts with the DNA. This doubling of contacts greatly increases the binding affinity while only doubling the free energy of the interaction. Helix-turn-helix motifs always bind to DNA that is in the B-DNA form.

The homeodomain motif is found in a special group of helix-turn-helix proteins that are encoded by homeotic selector genes, so called because the proteins encoded by these genes control developmental switches. For example, mutations in these genes cause one body part to be converted into another in the fruit fly, Drosophila. These genes have been found in every eukaryotic organism studied. The helix-turn-helix region of different homeodomains is always surrounded by the same structure, but not necessarily the same sequence, and the motif is always presented to DNA the same way. This helix-turn-helix configuration is stable by itself and, when isolated, can still bind to DNA. It may be significant that the helices in homeodomains are generally longer than the helices in most HLH regulatory proteins. Portions of the motif which interact most directly with DNA differ among these two families. Detailed examples of DNA-protein binding are described in Pabo and Sauer (1992; Ann Rev Biochem 61:1053-95).

A third motif incorporates zinc molecules into the crucial portion of the protein. These proteins are most often referred to as having zinc fingers, although their structure can be one of several types. Proteins in this family often contain tandem repeats of the 30-residue zinc finger motif, including the sequence patterns Cys-X_2 or 4-Cys-X_{12}-His-X_{3-5}-His. Each of these regulatory proteins has an α helix and an antiparallel ß sheet. Two histidines in the α helix and 2 cysteines near the turn in the ß sheet interact with the zinc ion which holds the α helix and the ß sheet together. Contact with the DNA is made by the arginine preceding the α helix, and by the second, third, and sixth residues of the α helix. When this arrangement is repeated as a cluster of several fingers, the α helix of each finger can contact and interact with the major groove of the DNA. By changing the number of zinc fingers, the specificity and strength of the binding interaction can be altered.

The steroid receptors are a family of intracellular proteins that include receptors for steroids, retinoids, vitamin D, thyroid hormones, and other important compounds. The DNA binding domain of these proteins contains about 70 residues, eight of which are conserved cysteines. The steroid receptor motif forms a structure in which two α helices are packed perpendicularly to each other, forming more of a globular shape than a finger. Each helix has a zinc ion which holds a peptide loop against the N-terminal end of the helix. The first helix fits into the major groove of DNA, and side chains make contacts with edges of the DNA base pairs. The steroid receptor proteins, like the helix-turn-helix proteins, form dimers that bind the DNA. The second helix of each monomer contacts the phosphate groups of the DNA backbone and also provides the dimerization interface. In some cases, multiple choices can exist for heterodimerization which produces another mechanism for fine-tuning the regulation of numerous genes.

Another family of regulatory protein molecules uses a motif consisting of a two-stranded antiparallel ß sheet to recognize the major groove of DNA. The exact DNA sequence recognized by the motif depends on the amino acid sequence in the ß sheet from which the amino acid side chains extend and contact the DNA. In two prokaryotic examples of the ß sheet, the regulatory proteins form tetramers when binding DNA.

The leucine zipper motif commonly forms dimers and has a 30-40 residue motif in which two α helices (one from each monomer) are joined to form a short coiled-coil. The helices are held together by interactions among hydrophobic amino acid side chains (often on heptad-repeated leucines) that extend from one side of each helix. Beyond this, the helices separate, and each basic region contacts the major groove of DNA. Proteins with the leucine zipper motif can also form either homodimers or heterodimers, thus extending the specific combinations available to activate or repress expression.

Yet another motif is the helix-loop-helix, which consists of a short α helix connected by a loop to a longer α helix. The loop is flexible and allows the two helices to fold back against each other. The α helices

bind both to DNA and to the HLH structure of another protein. The second protein can be the same (producing homodimers) or different (producing heterodimers). Some HLH monomers lack sufficient α helix to bind DNA, but they can still form heterodimers which can serve to inactivate specific regulatory proteins.

Hundreds of regulatory proteins have been identified to date, and more are being characterized in a wide variety of organisms. Most regulatory proteins have at least one of the common structural motifs for making contact with DNA, but several regulatory proteins, such as the p53 tumor suppressor gene, do not share their structure with other known regulatory proteins. Variations on the known motifs and new motifs have been and are currently being characterized (Faisst and Meyer (1992) Nucl Acids Res 20:3-26).

Although binding of DNA to a regulatory protein is very specific, there is no way to predict the exact DNA sequence to which a particular regulatory protein will bind or the primary structure of a regulatory protein for a specific DNA sequence. Thus, interactions of DNA and regulatory proteins are not limited to the motifs described above. Other domains of the proteins often form crucial contacts with the DNA, and accessory proteins can provide interactions which may convert a particular protein complex to an activator or a repressor or may prevent binding (Alberts et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York NY, pp.401-74).

Diseases and disorders related to gene regulation

Many neoplastic growths in humans can be traced to problems of gene regulation. Malignant growth of cells may be the result of excess transcriptional activator or loss of an inhibitor or suppressor (Cleary (1992) Cancer Surv 15:89-104). Alternatively, gene fusion may produce chimeric loci with switched domains, such that the level of activation is no longer correct for the gene specificity of that factor.

The cellular response to infection or trauma is beneficial when genes are appropriately expressed. However, when hyper-responsivity or another imbalance occurs for any reason, disregulation of gene expression may cause considerable tissue or organ damage. This damage is well documented in immunological responses to allergens, heart attack, stroke, and infections (Harrison's Principles of Internal Medicine, 13/e©, (1994) McGraw Hill and Teton Data Systems, Jackson WY). In addition, the accumulation of somatic mutations and the increasing inability to regulate cellular responses is seen in the prevalence of osteoarthritis and onset of other aging disorders.

The discovery of new human regulatory protein molecules which are expressed during disease development and the polynucleotides which encode them satisfies a need in the art by providing compositions which are useful in the diagnosis and treatment of diseases associated with cell proliferation, particularly immune responses and cancers.

SUMMARY OF THE INVENTION

The invention features purified proteins, human regulatory molecules, collectively referred to as HRM and individually referred to as HRM-1 through HRM-49. In one embodiment, the purified protein comprises an amino acid sequence selected from SEQ ID NO:1 through SEQ ID NO:49 and portions thereof..

The invention provides isolated polynucleotides encoding HRM and complements of the encoding polynucleotides. In one embodiment, the polynucleotide comprises a nucleic acid sequence selected from SEQ ID NOs:50-98 and complements thereof.

The invention also provides a polynucleotide, or a complement or a fragment thereof, which is used as a probe to hybridize to any one of the polynucleotides of SEQ ID NOs:50-98. The invention further provides a composition comprising the isolated and purified polynucleotides of SEQ ID NOs:50-98. In addition, the invention provides a composition comprising a polynucleotide selected from SEQ ID NOs:50-98 and complements and fragments thereof and a reporter molecule or stabilizing moiety. The invention still further provides a method for detecting expression of a polynucleotide which encodes a human regulatory molecule in a sample, the method comprising hybridizing the complement of a polynucleotide encoding HRM to nucleic acids of the sample under conditions to form a hybridization complex; and detecting hybridization complex formation, wherein complex formation indicates the expression of the polynucleotide encoding the human regulatory molecule in the sample. In one aspect, the complement of the polynucleotide encoding HRM is immobilized on a substrate. In another aspect, the substrate is a microarray.

The invention provides a vector containing at least a fragment of any one of the polynucleotides selected from SEQ ID NOs:50-98. In one embodiment, the vector is contained within a host cell. The invention also provides a method for producing a protein or a portion thereof, the method comprising culturing a host cell containing a vector containing at least a fragment of a polynucleotide encoding an HRM under conditions for the expression of the protein; and recovering the protein from the host cell culture.

The invention further provides a composition comprising a purified HRM and a labeling moiety or a pharmaceutical carrier. The invention still further provides a method for using an HRM to screen a plurality of molecules in order to obtain a ligand which specifically binds the HRM, the method comprising combining the protein with the molecules under conditions which allow specific binding, recovering the bound protein, separating the protein, thereby obtaining the ligand. In one aspect, the molecules are selected from libraries of agonists, antibodies, antagonists, drugs, inhibitors, peptides, proteins, and pharmaceutical agents.

The invention still further provides a method for using a protein to produce and purify an antibody, the method comprising immunizing a animal with an HRM under conditions to elicit an antibody response; isolating animal antibodies; attaching the protein to a substrate; contacting the substrate with sera containing antibodies under conditions to allow specific binding to the HRM; dissociating the antibodies from the HRM,

thereby obtaining purified antibodies.

The invention provides a purified antibody which specifically binds an HRM. The invention also provides a method for using an antibody to detect protein expression in a sample, the method comprising combining the antibody specifically binding HRM with a sample under conditions to form antibody:protein complexes and detecting complex formation, wherein detection indicates expression of the protein in the sample. In one aspect, expression of the HRM is diagnostic of cancer. In another aspect, expression is diagnostic of immune response.

The invention also provides a method for diagnosing a disease associated with gene expression in a sample containing nucleic acids, the method comprising hybridizing a polynucleotide to nucleic acids of the sample under conditions to form a hybridization complex, comparing hybridization complex formation to standards, thereby diagnosing the disease. In one aspect, the disease is selected from a disorder characterized by cell proliferation such as a cancer, an developmental disorder, or an immune response.

The invention provides a method for treating a cancer comprising administering to a subject in need of such treatment a composition containing purified HRM. The invention also provides a method for treating a cancer comprising administering to a subject in need of such treatment an antagonist which specifically binds HRM. The invention further provides a method for treating an immune response associated with the increased expression or activity of HRM comprising administering to a subject in need of such treatment an antagonist which specifically binds HRM. The invention still further provides a method for stimulating cell proliferation comprising administering purified HRM to a cell.

DESCRIPTION OF THE INVENTION

Before the present proteins, polynucleotide, and methods are described, it is to be understood that this invention is not limited to any particular methodology, protocol, cell line, vector, or reagent, as these may vary. It must also be noted that in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. For example, reference to "a host cell" includes a plurality of such host cells; reference to an "antibody" includes one or more antibodies and equivalents thereof known to those skilled in the art.

Unless defined herein, all technical and scientific terms have the same meanings commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology is used for the purpose of describing particular embodiments and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Although any methods and materials similar or equivalent to those described can be used in the practice or testing of the invention, the preferred methods, devices, and materials are now described. All publications are incorporated herein by reference for the purpose of

describing and disclosing the cell lines, vectors, arrays and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"Agonist" refers to a molecule which specifically binds to and modulates the activity of HRM.

An "allele" is an alternative form of the polynucleotide or gene encoding HRM. Alleles result from at least one mutation in the nucleic acid sequence and may result in the expression of altered mRNAs or proteins whose structure or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to additions, deletions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence. Similarly a polynucleotide may be altered to produce deliberate amino acid substitutions. Theses substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of HRM is retained. For example, negatively charged residues include aspartic acid and glutamic acid; positively charged residues include lysine and arginine; and residues with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Antagonist" refers to a molecule which, when bound to HRM, decreases the amount or the duration of the biological or immunological activity of HRM. Antagonists may include proteins, nucleic acids, carbohydrates, fats or any other molecules which decrease the effect of HRM.

"Antibody" refers to intact molecules, or fragments thereof such as Fa, $F(ab')_2$, and Fv, which are capable of binding the antigenic determinant of an HRM.

"Biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic protein or peptide to induce a specific immune response in animals or cells and to bind with specific antibodies.

"Complementary" refers to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". The degree of complementarity between nucleic acids has significant effects on the efficiency and strength of hybridization. This is important in amplification reactions and in the design and use of peptide nucleic acid molecules.

A "composition" refers to a combination comprising a plurality of polynucleotides or a specific

polynucleotide or protein and at least one other molecule. Such other molecules may include reporter molecules, labeling moieties, pharmaceutical carriers, carbohydrates, and the like.

"Consensus" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended using XL-PCR kit (Applied Biosystems (ABI), Foster City CA) in the 5' and/or the 3' direction and resequenced, or has been assembled to full length from overlapping shorter fragments using a computer program for fragment assembly such as that described in USSN 09/276,534, filed 25 March 1999.

"Derivative" refers to the chemical modification of a polynucleotide or protein. Such modifications may include replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative may encode a protein which retains the biological or immunological function of the natural molecule. A derivative protein is one which is modified by glycosylation, pegylation, or any similar process but still retains the biological or immunological function of the native protein.

"Differential expression" refers to an increased, upregulated or present, or decreased, downregulated or absent, gene expression as detected by presence, absence or at least about two-fold changes in the amount of transcribed messenger RNA or translated protein in a sample.

"Disorder" refers to a condition, disease or syndrome in which a polynucleotide or a protein of the invention is differentially expressed. Such a disorder includes cancers or immune responses as they are set forth below.

"HRM" refers to any one or all of the human proteins, HRMs 1-49, as it was obtained from any species including bovine, ovine, porcine, murine, equine, and preferably human, or from any source whether natural, synthetic, semi-synthetic, or recombinant.

"Hybridization complex" refers to a complex formed between two nucleic acids by the formation of hydrogen bonds between complementary base pairs; these hydrogen bonds form in an antiparallel configuration and may be further stabilized by base stacking interactions. A hybridization complex may be formed in solution or between one nucleic acid present in solution and another nucleic acid immobilized on a substrate.

"Isolated" refers to a polynucleotide that is removed from its natural environment or separated from other components with which it is naturally associated.

"Ligand" refers to any agent, molecule, or compound which will bind specifically to a polynucleotide or to a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic and/or organic substances including minerals, cofactors, nucleic acids, proteins, carbohydrates, fats, and lipids.

"Microarray" refers to an arrangement of distinct polynucleotides on a substrate

"Oligonucleotide" refers to a nucleic acid sequence about 6 nucleotides to about 60 nucleotides in length which may be used in amplification or hybridization assays. Equivalent terms include "amplimers","primers", "oligomers", and "probes", as these are commonly defined in the art.

"Peptide nucleic acid" refers to an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in a terminal lysine which confers solubility to the molecule.

"Polynucleotide" refers to nucleic acid molecule having a nucleic acid sequence and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded and represent the sense or antisense strand. "Fragment" refers to a nucleic acid sequence which is more than about 60 nucleotides in length.

"Portion" refers to a fragment of an HRM which ranges in size from five amino acid residues to the entire amino acid sequence minus one amino acid.

"Protein" refers to an oligopeptide, peptide, or polypeptide having an amino acid sequence whether naturally occurring or synthetic molecules. Portions of HRM are preferably about 5 to about 15 amino acids in length and retain the biological or the immunological activity of the HRM.

"Purified" refers to a peptide or protein that is removed from its natural environment, isolated or separated from other components with which it is naturally associated.

"Reporter molecules" or "labeling moieties" include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

"Sample" is used in its broadest sense and may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA (in solution or bound to a solid support), a tissue, a tissue print, and the like.

"Specific binding" refers to that interaction between a polynucleotide or protein of the invention and any ligand which specifically binds to it and which is selected from a DNA or an RNA molecule, a peptide nucleic acid, a peptide, a protein, an agonist, an antibody, an antagonist, an inhibitor, a mimetic, a pharmaceutical agent, a drug, a transcription factor, or an artificial chromosome construction. The interaction is dependent upon the presence of a particular sequence or three dimensional structure recognized by the binding molecule.

"Substrate" refers to any rigid or semi-rigid support to which polynucleotides or proteins are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

"Variant" refers to molecules that are recognized variations of a polynucleotide or a protein encoded by the polynucleotide. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to the polynucleotides and may differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a change in a single base as a result of a substitution, insertion or deletion. The change may be conservative (purine for purine) or non-conservative (purine to pyrimidine) and may or may not result in a change in an encoded amino acid or its secondary, tertiary, or quaternary structure.

THE INVENTION

The invention is based on the discovery of human regulatory molecules (HRM) and the polynucleotides encoding HRM, and on the use of these compositions for the diagnosis and treatment of diseases associated with cell proliferation. Table 1 shows the protein and polynucleotide identification numbers, protein abbreviation, Incyte Clone number, cDNA library, and the closest NCBI homolog and NCBI sequence identifier for each of the human regulatory molecules.

TABLE 1

Protein	Nucleotide	Abbreviation	Clone ID	Library	NCBI	Homolog
SEQ ID NO:1	SEQ ID NO:50	HRM-1	133	U937NOT01	g285947	KIAA0105
SEQ ID NO:2	SEQ ID NO:51	HRM-2	1762	U937NOT01	g1518121	Ascaris suum
SEQ ID NO:3	SEQ ID NO:52	HRM-3	1847	U937NOT01	g1302211	Saccharomyces cerevisiae
SEQ ID NO:4	SEQ ID NO:53	HRM-4	9337	HMC1NOT01	g1613852	Human zinc finger protein (zf2)
SEQ ID NO:5	SEQ ID NO:54	HRM-5	9476	HMC1NOT01	g755784	S. cerevisiae
SEQ ID NO:6	SEQ ID NO:55	HRM-6	10370	THP1PLB01	g895845	Human putative p64 CLCP protein
SEQ ID NO:7	SEQ ID NO:56	HRM-7	30137	THP1NOB01	g1710241	Human clone 23733 mRNA
SEQ ID NO:8	SEQ ID NO:57	HRM-8	77180	SYNORAB01	g5372	S. cerevisiae
SEQ ID NO:9	SEQ ID NO:58	HRM-9	98974	PITUNOR01	g1627704	Caenorhabditis elegans
SEQ ID NO:10	SEQ ID NO:59	HRM-10	118160	MUSCNOT01	g220594	Mus musculus
SEQ ID NO:11	SEQ ID NO:60	HRM-11	140516	TLYMNOR01	g1086723	C. elegans
SEQ ID NO:12	SEQ ID NO:61	HRM-12	207452	SPLNNOT02	g1314086	S. cerevisiae
SEQ ID NO:13	SEQ ID NO:62	HRM-13	208836	SPLNNOT02	g662126	S. cerevisiae
SEQ ID NO:14	SEQ ID NO:63	HRM-14	569710	MMLR3DT01	g1698719	Human zinc finger protein
SEQ ID NO:15	SEQ ID NO:64	HRM-15	606742	BRSTTUT01	g1710201	Human clone 23679 mRNA
SEQ ID NO:16	SEQ ID NO:65	HRM-16	611135	COLNNOT01	g506882	C elegans
SEQ ID NO:17	SEQ ID NO:66	HRM-17	641127	BRSTNOT03	g1310668	Human Hok-2 gene product
SEQ ID NO:18	SEQ ID NO:67	HRM-18	691768	LUNGTUT02	g309183	Mus musculus
SEQ ID NO:19	SEQ ID NO:68	HRM-19	724157	SYNOOAT01	g577542	C. elegans C16C10
SEQ ID NO:20	SEQ ID NO:69	HRM-20	864683	BRAITUT03	g1418563	C. elegans
SEQ ID NO:21	SEQ ID NO:70	HRM-21	933353	CERVNOT01	g1657672	C. elegans
SEQ ID NO:22	SEQ ID NO:71	HRM-22	1404643	LATRTUT02	g459002	C. elegans
SEQ ID NO:23	SEQ ID NO:72	HRM-23	1561587	SPLNNOT04	g868266	C. elegans
SEQ ID NO:24	SEQ ID NO:73	HRM-24	1568361	UTRSNOT05	g1834503	Human mucin
SEQ ID NO:25	SEQ ID NO:74	HRM-25	1572888	LNODNOT03	g603396	S. cerevisiae YER156c
SEQ ID NO:26	SEQ ID NO:75	HRM-26	1573677	LNODNOT03	g849195	S. cerevisiae D9481.16
SEQ ID NO:27	SEQ ID NO:76	HRM-27	1574624	LNODNOT03	g1067025	C. elegans R07E5.14
SEQ ID NO:28	SEQ ID NO:77	HRM-28	1577239	LNODNOT03	g728657	S. cerevisiae
SEQ ID NO:29	SEQ ID NO:78	HRM-29	1598203	BLADNOT03	g1200033	C. elegans F35G2

TABLE 1

SEQ ID NO:30	SEQ ID NO:79	HRM-30	1600438	BLADNOT03	g286001	KIAA0005
SEQ ID NO:31	SEQ ID NO:80	HRM-31	1600518	BLADNOT03	g790405	C. elegans
SEQ ID NO:32	SEQ ID NO:81	HRM-32	1602473	BLADNOT03	g1574570	Haemophilus influenzae
SEQ ID NO:33	SEQ ID NO:82	HRM-33	1605720	LUNGNOT15	g1055080	C. elegans
SEQ ID NO:34	SEQ ID NO:83	HRM-34	1610501	COLNTUT06	g313741	S. cerevisiae YBL0514
SEQ ID NO:35	SEQ ID NO:84	HRM-35	1720770	BLADNOT06	g1006641	C. elegans F46C5
SEQ ID NO:36	SEQ ID NO:85	HRM-36	1832295	BRAINON01	g561637	Human enigma protein
SEQ ID NO:37	SEQ ID NO:86	HRM-37	1990522	CORPNOT02	g558396	S. cerevisiae
SEQ ID NO:38	SEQ ID NO:87	HRM-38	2098087	BRAITUT02	g1066284	Mus musculus uterine mRNA
SEQ ID NO:39	SEQ ID NO:88	HRM-39	2112230	BRAITUT03	g861306	C. elegans
SEQ ID NO:40	SEQ ID NO:89	HRM-40	2117050	BRSTTUT02	g687821	C. elegans
SEQ ID NO:41	SEQ ID NO:90	HRM-41	2184712	SININOT01	g868241	C. elegans C56C10
SEQ ID NO:42	SEQ ID NO:91	HRM-42	2290475	BRAINON01	g733605	C. elegans
SEQ ID NO:43	SEQ ID NO:92	HRM-43	2353452	LUNGNOT20	g1507666	Schizosaccharomyces pombe
SEQ ID NO:44	SEQ ID NO:93	HRM-44	2469611	THP1NOT03	g1495332	C. elegans
SEQ ID NO:45	SEQ ID NO:94	HRM-45	2515476	LIVRTUT04	g1665790	KIAA0262
SEQ ID NO:46	SEQ ID NO:95	HRM-46	2754573	THP1AZS08	g478990	Human RNA binding protein
SEQ ID NO:47	SEQ ID NO:96	HRM-47	2926777	TLYMNOT04	g687823	C. elegans
SEQ ID NO:48	SEQ ID NO:97	HRM-48	3217567	TESTNOT07	g1841547	Human HLA class III region
SEQ ID NO:49	SEQ ID NO:98	HRM-49	3339274	SPLNNOT10	g1177434	Human mRNA

HRM-1 (SEQ ID NO:1) was identified in Incyte Clone 133 from the U937NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:50, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 133 (U937NOT01), 013508 (THP1PLB01), 210174 (SPLNNOT02), 1655863 (PROSTUT08), 1725724 (PROSNOT14), 1858205 (PROSNOT18), and 2646014 (OVARTUT05).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:1. HRM-1 is 151 amino acids in length and has four potential phosphorylation sites at T2, S14, S69, and T111. HRM-1 has sequence homology with human KIAA0105 (g285947) and is found in cDNA libraries which have proliferating cells and are associated with cancer or immune response.

HRM-2 (SEQ ID NO:2) was identified in Incyte Clone 1762 from the U937NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:51, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1762 (U937NOT01), 1254927 (LUNGFET03), and 2070865 (ISLTNOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:2. HRM-2 is 185 amino acids in length and has a potential N glycosylation site at N108; eight potential phosphorylation sites at T22, S26, T27, S31, T51, T70, and T135; a leucine zipper motif at L_{136}KDVVWGLNSLFTDLLNFDDPL; and a ubiquitin conjugation motif at W_{105}HPNITETGEICLSL. HRM-2 has sequence homology with a gene from Ascaris suum (g1518121) and is found in cDNA libraries which have secretory or proliferating cells and are associated with development.

HRM-3 (SEQ ID NO:3) was identified in Incyte Clone 1847 from the U937NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:52, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 274 (U937NOT01), 1847 (U937NOT01), 262233 (HNT2AGT01), 972977 (MUSCNOT02), and 1859611(PROSNOT18).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:3. HRM-3 is 59 amino acids in length and has four potential N glycosylation sites at N147, N352, N410, and N421, and 17 potential phosphorylation sites at S13, T21, S43, S89, S131, S207, T243, S278, T286, S335, S337, S350, S354, S369, S380, S412, and S542. HRM-3 has sequence homology with a Saccharomyces cerevisiae protein (g1302211) and is found in cDNA libraries which have proliferating or immortalized cells.

HRM-4 (SEQ ID NO:4) was identified in Incyte Clone 9337 from the HMC1NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:53, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 9337 (HMC1NOT01), 670279 (CRBLNOT01), 717305 (PROSTUT01), 968249 (BRSTNOT05), and 1546506 (PROSTUT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:4. HRM-4 is 338 amino acids in length and has a potential N glycosylation site at N327, 11 potential phosphorylation sites at T15, S36, S42, S50, T51, S73, S144, S176, T256, S140, and T329; and five zinc finger motifs at $C_{192}RC_{194}$SECGKI FRNPRYFSVHKKIH, C_{222}QDCGKGFVQSSSLTQHQRVH, C_{250}QECGRTFNDRSAISQHLRTH, C_{278}QDCGKAFRQSSHLIRHQRTH, and C_{306}NKCGKAFTQSSHLIGHQRTH. HRM-4 has sequence homology with a human zinc finger protein (g1613852) and is found in cDNA libraries which have proliferating, cancerous, or secretory cells.

HRM-5 (SEQ ID NO:5) was identified in Incyte Clone 9476 from the HMC1NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:54, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 9476 (HMC1NOT01), 010403 (THP1PLB01), 495099 (HNT2NOT01), 1670783 (BMARNOT03), 1997203 (BRSTTUT03), and 2190637 (THYRTUT03).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:5. HRM-5 is 456 amino acids in length and has a potential N glycosylation site at N385; 14 potential phosphorylation sites at T9, T12, S58, T74, T163, T139, S175, T211, T239, T272, S331, T367, T402, and S443; and an ATP/GTP binding motif at G_{70}PPGTGKT77. HRM-5 has sequence homology with a S. cerevisiae protein (g755784) and is found in cDNA libraries which have dividing, cancerous or immortalized cells and are associated with immune response.

HRM-6 (SEQ ID NO:6) was identified in Incyte Clone 10370 from the THP1PLB01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:55, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 010370 (THP1PLB01), 109018 (AMLBNOT01), 259388 (HNT2RAT01), and 1518624 (BLADTUT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:6. HRM-6 is 210 amino acids in length and has one potential N-glycosylation site at N11 and nine potential phosphorylation sites at T13, T21, T46, T124, S125, S132, T143, T167, and T191. HRM-6 has sequence homology with a putative p64 CLCP human protein (g895845) and is found in cDNA libraries which have dividing, cancerous or immortalized cells and are associated with immune response.

HRM-7 (SEQ ID NO:7) was identified in Incyte Clone 30137 from the THP1PLB01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:56, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 30137 (THP1NOB01), 531638 (BRAINOT03), 1653122 (PROSTUT08), and 1682227 (PROSNOT15).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ

ID NO:7. HRM-7 is 255 amino acids in length and has one potential N glycosylation site at N86 and 12 potential phosphorylation sites at T9, T28, S32, S61, S94, S142, S156, S160, T169, S188, S220, and S236. HRM-7 has sequence homology with human clone 23733 (g1710241) and is found in cDNA libraries which have dividing, cancerous or immortalized cells and are associated with immune response.

HRM-8 (SEQ ID NO:8) was identified in Incyte Clone 77180 from the SYNORAB01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:57, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 077180 (SYNORAB01), 604706 (BRSTTUT01), 977901 (BRSTNOT02), 1870373 (SKINBIT01), and 2169441 (ENDCNOT03).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:8. HRM-8 is 188 amino acids in length and has one potential amidation site, Q170GKR; two potential N glycosylation sites at N60 and N68; and four potential phosphorylation sites at S70, T164, T166, and S183. HRM-8 has sequence homology with a S. cerevisiae protein (g5372) and is found in cDNA libraries which have dividing, cancerous or immortalized cells and are associated with immune response.

HRM-9 (SEQ ID NO:9) was identified in Incyte Clone 98974 from the PITUNOR01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:58, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 98974 (PITUNOR01), 443924 (MPHGNOT03), 1401540 (BRAITUT08), 1507305 (BRAITUT07), 1700814 (BLADTUT05), and 1809947 (PROSTUT12).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:9. HRM-9 is 531 amino acids in length and has one potential N glycosylation site at N480; 37 potential phosphorylation sites at S19, T22, S38, T64, T76, T91, S117, S118, S158, T164, T177, T182, T200, T267, Y281, Y311, Y322, S333, S394, S402, S404, S409, S414, S416, S418, S429, S434, S439, S440, S456, S460, S466, S478, S505, S510, S524, S528, and one potential glycosaminoglycan motif at S434GSG. HRM-9 has sequence homology with a Caenorhabditis elegans protein (g1627704) and is found in cDNA libraries which have secretory, proliferating or immune cells.

HRM-10 (SEQ ID NO:10) was identified in Incyte Clone 118160 from the MUSCNOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:59, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 118160 (MUSCNOT01), 323015 (EOSIHET02), and 1856519 (PROSNOT18).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:10. HRM-10 is 348 amino acids in length and has two potential N glycosylation sites at N150 and N317; 17 potential phosphorylation sites at T23, T45, S60, T126, S130, S140, S145, S151, S154, S158, S186,

Y208, Y234, S217, T271, T303, and S327, and a transcription factor signature at C_{310}SKCKKKNCTYNQVQTRSA DEPMTTFVLCNEC. HRM-10 has sequence homology with a Mus musculus protein (g220594) and is found in cDNA libraries which have secretory or immune associations.

HRM-11 (SEQ ID NO:11) was identified in Incyte Clone 140516 from the TLYMNOR01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:60, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 140516 (TLYMNOR01), 143729 (TLYMNOR01), 1346014 (PROSNOT11), and 2074866 (ISLTNOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:11. HRM-11 is 393 amino acids in length and has 14 potential phosphorylation sites at S22, T33, S41, S69, T156, Y157, S166, S199, T242, T308, T324, S350, T359, S378. HRM-11 has sequence homology with a C. elegans protein (g1086723) and is found in cDNA libraries which have proliferating, secretory or immune cells.

HRM-12 (SEQ ID NO:12) was identified in Incyte Clone 207452 from the SPLNNOT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:61, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 207452 (SPLNNOT02), 238306 (SINTNOT02), 1559492 (SPLNNOT04), and 1852567 (LUNGFET03).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:12. HRM-12 is 320 amino acids in length and one potential amidation site at E_{210}GKK; two potential N glycosylation sites at N12 and N314; seven potential phosphorylation sites at S34, S51, S56, S111, T157, S198, and S318; one potential glycosaminoglycan motif, S224GAG; one immunoglobulin major histocompatibility motif, F_{305}FCNVFH; and two mitochondrial carrier protein signatures, P_{35}FDVIKIRF and P_{138}VDVLRTRF. HRM-12 has sequence homology with a S. cerevisiae protein (g1314086) and is found in cDNA libraries which have secretory and proliferating cells.

HRM-13 (SEQ ID NO:13) was identified in Incyte Clone 208836 from the SPLNNOT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:62, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 26879 (SPLNFET01), 208836 (SPLNNOT02), and 1916142 (PROSTUT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:13. HRM-13 is 343 amino acids in length and has one potential N glycosylation site at N172; 17 potential phosphorylation sites at S45, S46, T62, S73, S84, S85, S102, S105, T124, S137, Y153, T192, S216, Y226, Y241, S253 and T293; and a zinc finger motif at C_{277}RHYFCESCA. HRM-13 has sequence homology with a S. cerevisiae protein (g662126) and is found in cDNA libraries which have proliferating cells and are

associated with immune response.

HRM-14 (SEQ ID NO:14) was identified in Incyte Clone 569710 from the MMLR3DT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:63, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 145344 (TLYMNOR01) and 569710 (MMLR3DT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:14. HRM-14 is 368 amino acids in length and has 10 potential phosphorylation sites at S5, T16, T125, S132, S142, S157, S167, S185, S208, and S246; and four zinc finger motifs at C_{253}DECGKHFSQGSALILHQRIH, C_{281}VECGKAFSRSSILVQH QRVH, C_{309}LECGKAFSQNSGLINHQRIH, and C_{337}VQCGKSYSQSSNLFRHQRRH. HRM-14 has sequence homology with a human zinc finger protein (g1698719) and is found in cDNA libraries which are associated with immune response.

HRM-15 (SEQ ID NO:15) was identified in Incyte Clone 606742 from the BRSTTUT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:64, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 606742 (BRSTTUT01) and 1559478 (SPLNNOT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:15. HRM-15 is 158 amino acids in length and has two potential myristylation sites, G92GFHGQ and G96QMHSR, and one potential PKC phosphosphorylation site, S40. HRM-15 has sequence homology with human clone 23679 (g1710201) and is found in cDNA libraries with proliferating, secretory and/or cancerous cells.

HRM-16 (SEQ ID NO:16) was identified in Incyte Clone 611135 from the COLNNOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:65, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 611135 (COLNNOT01), 659029 (BRAINOT03), and 1861691 (PROSNOT19).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:16. HRM-16 is 334 amino acids in length and has 11 potential phosphorylation sites at S17, T29, T128, S133, S162, S176, S263, T257, S263, S277, and S294. HRM-16 has sequence homology with a C. elegans protein (g506882) and is found in cDNA libraries with secretory cells.

HRM-17 (SEQ ID NO:17) was identified in Incyte Clone 641127 from the BRSTNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:66, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 641127

(BRSTNOT03) and 673153 (CRBLNOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:17. HRM-17 is 488 amino acids in length and has one N glycosylation site at N215; 11 potential phosphorylation sites at S70, S78, S92, T102, S111, T190, Y235, S303, S329, S415, and T471; and eight zinc finger motifs at C_{237}EQCGKGFTRSSSLLIHQAVH, C_{265}DKCGKGFTRSSSLLIHHAVH, C_{293}DKCGKGFSQSSKLHIHQRVH, C_{321}EECGMSFS QRSNLHIHQRVH , C_{349}GECGKGFSQSSNLHIHRCIH, C_{377}YECGKGFSQSSDLRIHLRVH, C_{405}GKCGKGFSQSSKLLIHQRVH, and C_{433}SKCGKGFSQSSNLHIHQRVH. HRM-17 has sequence homology with a human HOK-2 gene product (g1310668) and is found in cDNA libraries associated with sensory and secretory functions.

HRM-18 (SEQ ID NO:18) was identified in Incyte Clone 691768 from the LUNGTUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:67, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 691768 (LUNGTUT02), 1417161 (BRAINOT12) and 1931861 (COLNNOT16).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:18. HRM-18 is 255 amino acids in length and has one potential N glycosylation site at N102 and 13 potential phosphorylation sites at S21, T90, T109, S111, T124, S134, S139, T141, S158, S172, S181, S187, and T206. HRM-18 has sequence homology with a M. musculus protein (g309183) and is found in cDNA libraries with proliferating or cancerous cells.

HRM-19 (SEQ ID NO:19) was identified in Incyte Clone 724157 from the SYNOOAT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:68, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 724157 (SYNOOAT01), 1516153 (PANCTUT01), and 1610152 (COLNTUT06).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:19. HRM-19 is 351 amino acids in length and has eight potential phosphorylation sites at T30, S41, S53, T135, S172, S187, T273, and S331; one potential glycosaminoglycan site, S_{18}GTG; and one potential mitochondrial carrier motif, P_{31}LDVVKVRL. HRM-19 has sequence homology with C. elegans C16C10 (g577542) and is found in cDNA libraries associated with cell proliferation, cancer and immune response.

HRM-20 (SEQ ID NO:20) was identified in Incyte Clone 864683 from the BRAITUT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:69, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 486297 (HNT2RAT01), 864683 (BRAITUT03), 1314465 (BLADTUT02), 1610776 (COLNTUT06), 1856771

(PROSNOT18), 1866081 (PROSNOT19), 1932221 (COLNNOT16), and 2125225 (BRSTNOT07).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:20. HRM-20 is 535 amino acids in length and has three potential N glycosylation sites at N202, N252, and N523; and 17 potential phosphorylation sites at S2, S12, S42, S49, S102, S157, T165, T171, T232, T255, T317, S332, S428, T441, S453, S500, and S509. HRM-20 has sequence homology with a C. elegans protein (g1418563) and is found in cDNA libraries associated with cell proliferation, cancer and immune response.

HRM-21 (SEQ ID NO:21) was identified in Incyte Clone 933353 from the CERVNOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:70, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 928904 (BRAINOT04), 933353 (CERVNOT01), and 2452674 (ENDANOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:21. HRM-21 is 201 amino acids in length and has one potential N glycosylation site at N82; five potential phosphorylation sites at T70, S83, S98, S154, and T187; and one tyrosine phosphatase motif at V_{130}HCKAGRSRSATM. HRM-21 has sequence homology with a C. elegans protein (g1657672) and is found in cDNA libraries associated with immune response.

HRM-22 (SEQ ID NO:22) was identified in Incyte Clone 1404643 from the LATRTUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:71, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 878243 (LUNGAST01), 1404643 (LATRTUT02), 1508343 (LUNGNOT14) and 2585156 (BRAITUT22).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:22. HRM-22 is 239 amino acids in length and has four potential phosphorylation sites at S5, S89, S133, and T211. HRM-22 has sequence homology with a C. elegans protein (g459002) and is found in cDNA libraries associated with cell proliferation, cancer and immune response.

HRM-23 (SEQ ID NO:23) was identified in Incyte Clone 1561587 from the SPLNNOT04 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:72, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 522573 (MMLR2DT01), 773822 (COLNNOT05), 1304839 (PLACNOT02), 1381253 (BRAITUT08), 1452511 (PENITUT01), 1539060 (SINTTUT01), 1561587 (SPLNNOT04), and 2416572 (HNT3AZT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:23. HRM-23 is 244 amino acids in length and has five potential phosphorylation sites at T40, S75, T84, T89, and S194. HRM-23 has sequence homology with a C. elegans protein (g868266) and is found in cDNA libraries associated with cell proliferation, cancer and immune response.

HRM-24 (SEQ ID NO:24) was identified in Incyte Clone 1568361 from the UTRSNOT05 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:73, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 927874 (BRAINOT04), 1255220 (MENITUT03), 1242340 (LUNGNOT03), 1349495 (LATRTUT02), 1381263 (BRAITUT08), 1500028 (SINTBST01), 1568361 (UTRSNOT05), 1653237 (PROSTUT08), 1975340 (PANCTUT02), and 3274608 (PROSBPT06).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:24. HRM-24 is 431 amino acids in length and has five potential N glycosylation sites at N75, N95, N171, N202, and N298; eight potential phosphorylation sites at S2, S3, T11, T13, S17, Y316, T375, and T415; and a leucine zipper motif, L_{96}SAFNNILSNLGYILLGLLFLL. HRM-24 has sequence homology with human mucin (g1834503) and is found in cDNA libraries with proliferating, cancerous or inflamed cells.

HRM-25 (SEQ ID NO:25) was identified in Incyte Clone 1572888 from the LNODNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:74, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1438142 (PANCNOT08), 1572888 (LNODNOT03), and 1665075 (BRSTNOT09).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:25. HRM-25 is 376 amino acids in length and has one N glycosylation site at N51 and five potential phosphorylation sites at S111, T150, S151, T159, and S196. HRM-25 has sequence homology with S. cerevisiae YER156c (g603396) and is found in cDNA libraries with secretory cells.

HRM-26 (SEQ ID NO:26) was identified in Incyte Clone 1573677 from the LNODNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:75, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 040360 (TBLYNOT01), 065573 (PLACNOB01), 228382 (PANCNOT01), 1456688 (COLNFET02), 1573677 (LNODNOT03), and1854560 (HNT3AZT01)

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:26. HRM-26 is 340 amino acids in length and has one potential N glycosylation site at N213 and 13 potential phosphorylation sites at T10, S22, T53, T56, S160, S168, S170, S177, S201, S226, S297, S303, and T329. HRM-26 has sequence homology with S. cerevisiae D9481.16 (g849195) and is found in cDNA libraries associated with secretion, immune response, and cancer.

HRM-27 (SEQ ID NO:27) was identified in Incyte Clone 1574624 from the LNODNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:76, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 90012

(HYPONOB01), 888491 (STOMTUT01), and 1574624 (LNODNOT03).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:27. HRM-27 is 174 amino acids in length and has one N glycosylation site at N51 and five potential phosphorylation sites at S111, T150, S151, T159, and S196. HRM-27 has sequence homology with a C. elegans protein (g1067025) and is found in cDNA libraries associated with secretion, immune response, and cancer.

HRM-28 (SEQ ID NO:28) was identified in Incyte Clone 1577239 from the LNODNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:77, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 100565 (ADRENOT01), 1336693 (COLNNOT13), and 1577239 (LNODNOT03).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:28. HRM-28 is 179 amino acids in length and has one potential N glycosylation site at N60 and five potential phosphorylation sites at Y61, S62, Y104, T136, and Y142. HRM-28 has sequence homology with a S. cerevisiae protein (g728657) and is found in cDNA libraries associated with secretion and immune response.

HRM-29 (SEQ ID NO:29) was identified in Incyte Clone 1598203 from the BLADNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:78, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1598203 (BLADNOT03), 1697035 (COLNNOT23), and 1932332 (COLNNOT16).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:29. HRM-29 is 205 amino acids in length and has one potential N glycosylation site at N117 and five potential phosphorylation sites at T68, T118, S137, S140, and S159. HRM-29 has sequence homology with a C. elegans protein (g1200033) and is found in cDNA libraries associated with secretion.

HRM-30 (SEQ ID NO:30) was identified in Incyte Clone 1600438 from the BLADNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:79, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 835283 (PROSNOT07), 1600044 (BLADNOT03), 1600438 (BLADNOT03), and 1922072 (BRSTTUT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:30. HRM-30 is 419 amino acids in length and has one potential N glycosylation site at N161; twelve potential phosphorylation sites at T16, S57, T67, T83, S100, T107, S144, S206, T254, Y351,S412, and S414; a leucine zipper motif, L_{38}NEAGDDLEAVAKFLDSTGSRL; and an ATP/GTP binding motif, A_{385}HVAKGKS. HRM-30 has sequence homology with human KIAA0005 (g286001) and is found in

cDNA libraries associated with secretion and cancer.

HRM-31 (SEQ ID NO:31) was identified in Incyte Clone 1600518 from the BLADNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:80, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 389679 (THYMNOT02), 1600518 (BLADNOT03), 2055734 (BEPINOT01), 2102793 (BRAITUT02), and 2509270 (CONUTUT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:31. HRM-31 is 376 amino acids in length and has one potential N glycosylation site at N161 and 14 potential phosphorylation sites at T30, S65, S75, S95, S106, T134, S159, S224, T228, T250, T292, S299 ,T303, and S323 and a glycosaminoglycan motif, S14GPG. HRM-31 has sequence homology with a C. elegans protein (g790405) and is found in cDNA libraries associated with immune response, secretion, and cancer.

HRM-32 (SEQ ID NO:32) was identified in Incyte Clone 1602473 from the BLADNOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:81, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1351857 (LATRTUT02), 1602473 (BLADNOT03), and 2478778 (SMCANOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:32. HRM-32 is 237 amino acids in length and has seven potential phosphorylation sites at T51, T68, S92, S143, T171, S193, and S203. HRM-32 has sequence homology with a Haemophilus influenzae protein (g1574570) and is found in cDNA libraries associated with immune response, secretion, and cancer.

HRM-33 (SEQ ID NO:33) was identified in Incyte Clone 1605720 from the LUNGNOT15 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:82, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 660915 (BRAINOT03), 1347135 (PROSNOT11), and 1605720 (LUNGNOT15).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:33. HRM-33 is 152 amino acids in length and has four potential phosphorylation sites at S10, S23, T34, and S66; and a leucine zipper motif, L_{77}AVGNYRLKEYEKALKYVRGLL. HRM-33 has sequence homology with C. elegans (g1055080) and is found in cDNA libraries associated with secretion and immune response.

HRM-34 (SEQ ID NO:34) was identified in Incyte Clone 1610501 from the COLNTUT06 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:83, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1610501 (COLNTUT06) and 2477716 (SMCANOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:34. HRM-34 is 179 amino acids in length and has five potential phosphorylation sites at S32, S48, T45, T50, and T52. HRM-34 has sequence homology with a S. cerevisiae protein (g313741) and is found in cDNA libraries associated with cancer and immune response.

HRM-35 (SEQ ID NO:35) was identified in Incyte Clone 1720770 from the BLADNOT06 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:84, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 681455 (UTRSNOT02), 813292 (LUNGNOT04), 1223029 (COLNTUT02), 1444186 (THYRNOT03), 1522592 (BLADTUT04), 1720770 (BLADNOT06), and 1798409 (COLNNOT27).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:35. HRM-35 is 196 amino acids in length and has an amidation motif, $H_{17}9GKR$, and seven potential phosphorylation sites at S2, S6, S31,S84, S90, T136, and T161. HRM-35 has sequence homology with a C. elegans protein (g1006641) and is found in cDNA libraries associated with secretion, immune response, and cancer.

HRM-36 (SEQ ID NO:36) was identified in Incyte Clone 1832295 from the BRAINON01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:85, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 060275 (LUNGNOT01), 1823989 (GBLATUT01), and 1832295 (BRAINON01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:36. HRM-36 is 612 amino acids in length and has 12 potential N glycosylation sites at N36, N95, N139, N146, N151, N176, N188, N226, N243, N353, N371, and N482; and 16 potential phosphorylation sites at S58, S92, S112, T153, T198, T248, S308, S373, T400, T420, T428, Y438, T458, T472, S527, and S556. HRM-36 has sequence homology with human enigma protein (g561637) and is found in cDNA libraries associated with secretion and immune response.

HRM-37 (SEQ ID NO:37) was identified in Incyte Clone 1990522 from the CORPNOT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:86, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 264363 (HNT2AGT01), 1990522 (CORPNOT02), and 2451448 (ENDANOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:37. HRM-37 is 101 amino acids in length and has a PKC phosphorylation site at S62. HRM-37 has sequence homology with a S. cerevisiae protein (g558396) and is found in cDNA libraries associated with immune response.

HRM-38 (SEQ ID NO:38) was identified in Incyte Clone 2098087 from the BRAITUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:87, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 690359 (LUNGTUT02), 1429907 (SINTBST01), and 2098087 (BRAITUT02).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:38. HRM-38 is 132 amino acids in length and has a potential ATP/GTP binding motif at G_{74}ARNLLKS. HRM-38 has sequence homology with M. musculus uterine protein (g1066284) and is found in cDNA libraries associated with immune response.

HRM-39 (SEQ ID NO:39) was identified in Incyte Clone 2112230 from the BRAITUT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:88, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1383278 (BRAITUT08), 1646103 (PROSTUT09), 2112230 (BRAITUT03), and 2510591 (CONUTUT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:39. HRM-39 is 188 amino acids in length and has a potential N glycosylation site at N87 and eight potential phosphorylation sites at T10, T28, S74, S93, T121, T128, Y168, and T169. HRM-39 has sequence homology with a C. elegans protein (g861306) and is found in cDNA libraries from cancerous tissues.

HRM-40 (SEQ ID NO:40) was identified in Incyte Clone 2117050 from the BRAITUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:89, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 941515 (ADRENOT03), 1549443 (PROSNOT06), 2113261 (BRAITUT03), 2117050 (BRSTTUT02), and 2530536 (GBLANOT02).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:40. HRM-40 is 86 amino acids in length and has a potential N glycosylation site at N58 and four potential phosphorylation sites at T2, S9, T26, and T27. HRM-40 has sequence homology with a C. elegans protein (g687821) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-41 (SEQ ID NO:41) was identified in Incyte Clone 2184712 from the SININOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:90, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 922736 (RATRNOT02), 1976003 (PANCTUT02), and 2184712 (SININOT01).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:41. HRM-41 is 222 amino acids in length and has a potential amidation site, K_{10}GKK; a potential

glycosaminoglycan site, S_2GLG; a potential N glycosylation site, N95; and seven potential phosphorylation sites at T18, T29, T50, S84, T98, S112, and S188. HRM-41 has sequence homology with a C. elegans protein (g868241) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-42 (SEQ ID NO:42) was identified in Incyte Clone 2290475 from the BRAINON01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:91, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 238339 (SINTNOT02), 1657945 (URETTUT01), 1848691 (LUNGFET03), 2044604 (THP1T7T01), 2290475 (BRAINON01), and 2514944 (LIVRTUT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:42. HRM-42 is 300 amino acids in length and has a potential N glycosylation site, N5; seven potential phosphorylation sites at S23, S71, S132, S142, T176, T192, and S293; and a Mutt signature, G_{165}MVDPGEKISATLKREFGEE. HRM-42 has sequence homology with a C. elegans protein (g733605) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-43 (SEQ ID NO:43) was identified in Incyte Clone 2353452 from the LUNGNOT20 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:92, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1000164 (BRSTNOT03), 1308080 (COLNFET02), 1900151 (BLADTUT06), and 2353452 (LUNGNOT20).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:43. HRM-43 is 112 amino acids in length and has six potential phosphorylation sites at T23, T43, S44, T79, T84, and T98. HRM-43 has sequence homology with a Schizosaccharomyces pombe protein (g1507666) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-44 (SEQ ID NO:44) was identified in Incyte Clone 2469611 from the THP1NOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:93, was derived from the extended and overlapping nucleic acid sequences: Incyte clones 003088 (HMC1NOT01), 1448981 (PLACNOT02), 1453563 (PENITUT01), 1824146 (GBLATUT01), 2369282 (ADRENOT07), 2469611 (THP1NOT03), and 2622587 (KERANOT02).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:44. HRM-44 is 251 amino acids in length and has a potential glycosaminoglycan site, S218GFG, and four potential phosphorylation sites at T8, S83, S212, and S226. HRM-44 has sequence homology with a C. elegans protein (g1495332) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and

immune response.

HRM-45 (SEQ ID NO:45) was identified in Incyte Clone 2515476 from the LIVRTUT04 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:94, was derived from the extended and overlapping nucleic acid sequences: Incyte clones 18414 (HUVELPB01), 78341 (SYNORAB01), 143277 (TLYMNOR01), 181574 (PLACNOB01), 832996 (PROSTUT04), 962753 (BRSTTUT03), 1413604 (BRAINOT12), and 2515476 (LIVRTUT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:45. HRM-45 is 811 amino acids in length and has three potential amidation sites at G_{113}GRR, W_{165}GKR, and G_{790}GKK; four potential N glycosylation sites at N22, N56, N79, and N145; 24 potential phosphorylation sites at T11, S13, S30, S60, Y71, S81, S85, S86, S103, S254, S256, T377, S388, S425, S456, S487, T544, S552, S574, T659, S678, S702, S746, and S753; a potential glycosaminoglycan site, S_{160}GHG; and a potential zinc finger motif at C_{240}GHIFCWACI. HRM-45 has sequence homology with human KIAA0262 (g1665790) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-46 (SEQ ID NO:46) was identified in Incyte Clone 2754573 from the THP1AZS08 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:95, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 263630 (HNT2AGT01), 412307 (BRSTNOT01), 491644 (HNT2AGT01), 1253094 (LUNGFET03), 2270603 (PROSNON01), 2280508 (PROSNON01), 2375670 (ISLTNOT01), 2754573 (THP1AZS08), and 3151587 (ADRENON04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:46. HRM-46 is 352 amino acids in length and has two potential N glycosylation sites at N141 and N294, and thirteen potential phosphorylation sites at S8, T67, T106, T110, T121, S122, S169, S206, T210, S215, S256, S260, and T296. HRM-46 has sequence homology with human RNA binding protein (g478990) and is found in cDNA libraries involved in cell proliferation, secretion, and immune response.

HRM-47 (SEQ ID NO:47) was identified in Incyte Clone 2926777 from the TLYMNOT04 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:96, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 040208 (TBLYNOT01), 900242 (BRSTTUT03), 963500 (BRSTTUT03), 1996474 (BRSTTUT03), and 2926777 (TLYMNOT04).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:47. HRM-47 is 432 amino acids in length and has a potential N glycosylation site at N417 and 24

potential phosphorylation sites at T51, S73, T122, T133, S177, S206, T226, T238, S293, S300, S304, S309, T325, S333, S339, S353, S360, Y361, S384, S390, T403, T412, T419, and S425 . HRM-47 has sequence homology with a C. elegans protein (g687823) and is found in cDNA libraries involved in cell proliferation, secretion, cancer, and immune response.

HRM-48 (SEQ ID NO:48) was identified in Incyte Clone 3217567 from the TESTNOT07 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:97, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 905037 (COLNNOT07), 1287503 (BRAINOT11), and 3217567 (TESTNOT07).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:48. HRM-48 is 180 amino acids in length and has a potential zinc finger motif, C42GHLYCWPCL, and five potential phosphorylation sites at T33, T57, S84, T148, and S160. HRM-48 has sequence homology with human HLA class III region (g1841547) and is found in cDNA libraries involved in secretion and immune response.

HRM-49 (SEQ ID NO:49) was identified in Incyte Clone 3339274 from the SPLNNOT10 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:98, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 532254 (BRAINOT03), 941336 (ADRENOT03), 2447649 (THP1NOT03), and 3339274 (SPLNNOT10).

In one embodiment, the invention encompasses a protein comprising the amino acid sequence of SEQ ID NO:49. HRM-49 is 137 amino acids in length and has three potential phosphorylation sites at T11, T91, and S119. HRM-49 has sequence homology with a deduced human translational inhibitor (g1177434) and is found in cDNA libraries involved in secretion and immune response.

The invention also encompasses HRM variants which retain the biological or functional activity of HRM. A preferred HRM variant is one having at least 60% amino acid sequence identity to an amino acid sequence selected from SEQ ID NOs:1-49.

The invention also encompasses polynucleotides which encode HRM. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of HRM can be used to produce recombinant molecules which express HRM. In a particular embodiment, the invention encompasses a polynucleotide comprising a nucleic acid sequence selected from SEQ ID NOs:50-98 and fragment and complements thereof.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotides encoding HRM, some bearing minimal homology to the polynucleotides of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide that could be made by selecting combinations based on possible codon

choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide of naturally occurring HRM, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode HRM and its variants are preferably capable of hybridizing to the polynucleotide of the naturally occurring HRM under selected conditions of stringency, it may be advantageous to produce polynucleotides encoding HRM or its derivatives possessing a different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for altering the polynucleotide encoding HRM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides, or fragments thereof, which encode HRM and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HRM or any fragment thereof.

Also encompassed by the invention are polynucleotides that are capable of hybridizing to the nucleic acids of a sample, and in particular, the polynucleotides or the complements thereof shown in SEQ ID NOs:50-98, under various conditions of stringency as taught in Wahl and Berger (1987; Methods Enzymol 152:399-407) and Kimmel (1987; Methods Enzymol 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, the process is automated with machines such as the MICROLAB system (Hamilton, Reno NV), DNA ENGINE thermal cycler (MJ Research, Watertown MA), and the Catalyst preparation and 373 and 377 PRISM DNA sequencing systems (ABI).

The nucleic acid sequences encoding HRM may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar (1993) PCR Methods Applic

2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia et al. (1988) Nucleic Acids Res 16:8186). The primers may be designed using commercially available software such as OLIGO software (Molecular Insights, Cascade CO), or another program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72 C. The method uses several restriction enzymes to generate a fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom et al. (1991) PCR Methods Applic 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker et al. (1991; Nucleic Acids Res 19:3055-3060). One may also use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This process avoids the need to screen libraries of cDNAs for longer sequences and is very useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using software integral to the system, and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be

present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode HRM may be used in recombinant DNA molecules to direct expression of HRM, portions or functional equivalents thereof, in host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express HRM.

As will be understood by those of skill in the art, it may be advantageous to produce HRM-encoding polynucleotides possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The polynucleotides of the present invention can be engineered using methods generally known in the art in order to alter HRM encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the polynucleotides. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HRM may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of HRM activity, it may be useful to encode a chimeric HRM protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the HRM encoding sequence and the heterologous protein sequence, so that HRM may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding HRM may be synthesized, in whole or in part, using chemical methods well known in the art (Caruthers et al. (1980) Nucleic Acids Symp Ser. (7) 215-223, Horn et al. (1980) Nucleic Acids Symp. Ser. (7) 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of HRM, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the 431A Peptide synthesizer (ABI).

The newly synthesized peptide may be purified by preparative high performance liquid chromatography (see Creighton (1983) Proteins, Structures and Molecular Principles, WH Freeman, New York NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or

sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of HRM, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant protein.

In order to express a biologically active HRM, the polynucleotides encoding HRM or functional equivalents, may be inserted into expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HRM and transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook et al. (1989; Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY) and Ausubel et al. (1989; Current Protocols in Molecular Biology, John Wiley & Sons, New York NY).

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HRM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with baculovirus expression vectors; plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus or tobacco mosaic virus) or with bacterial expression vectors (Ti or pBR322 plasmids); or animal cell systems.
The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla CA) or the pSport1 plasmid (Life Technologies) may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector if the protein is to be produced in plant cells. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding HRM, vectors based on SV40 or EBV may be used with an selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for HRM. For example, when large quantities of HRM are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as BLUESCRIPT phagemid (Stratagene), in which the sequence encoding HRM may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke and Schuster (1989) J Biol Chem 264:5503-5509); and the like. pGEX vectors (APB) may also be used to express foreign proteins as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned protein of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel (supra) and Grant et al. (1987; Methods Enzymol 153:516-544).

In cases where plant cell expression is desired, the expression of sequences encoding HRM may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu (1987) EMBO J 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi et al. (1984) EMBO J 3:1671-1680; Broglie et al. (1984) Science 224:838-843; and Winter et al. (1991) Results Probl Cell Differ 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs or Murry, In: McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY; pp. 191-196).

An insect system may also be used to express HRM. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding HRM may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of HRM will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which HRM may be expressed (Engelhard et al. (1994) Proc Nat Acad Sci 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HRM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing HRM in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HRM. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding HRM, its initiation codon, and upstream sequences are inserted into the expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are for the particular cell system which is used, such as those described in the literature (Scharf et al. (1994) Results Probl Cell Differ 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the ATCC (Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express HRM may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for

1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy et al. (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler et al. (1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin et al (1981) J Mol Biol 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan (1988) Proc Natl Acad Sci 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes et al. (1995) Methods Mol Biol 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding HRM is inserted within a marker gene sequence, transformed cells containing sequences encoding HRM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HRM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding HRM and express HRM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotides encoding HRM can be detected by DNA-DNA or DNA-RNA hybridization or PCR amplification. Nucleic acid amplification based assays involve the use of oligonucleotides based on the polynucleotides encoding HRM to detect transformants containing DNA or

RNA encoding HRM.

A variety of protocols for detecting and measuring the expression of HRM, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HRM is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox et al. (1983; J Exp Med 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HRM include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HRM, or any fragments thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (APB; Promega, Madison WI).

Host cells transformed with polynucleotides encoding HRM may be cultured under conditions for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HRM may be designed to contain signal sequences which direct secretion of HRM through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding HRM to polynucleotide encoding a protein domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex, Seattle WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and HRM may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing HRM and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography) as described in Porath et al. (1992, Prot Exp Purif 3:263-281) while the enterokinase cleavage site provides a means for purifying HRM from the

fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll et al. (1993; DNA Cell Biol 12:441-453).

In addition to recombinant production, portions of HRM may be produced by direct peptide synthesis using solid-phase techniques (Merrifield (1963) J Am Chem Soc 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using 431A Peptide synthesizer (ABI). Various portions of HRM may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exits among the human regulatory proteins of the invention. The expression of HRM is closely associated with cell proliferation. Therefore, in cancers or immune disorders where HRM is an activator, transcription factor, or enhancer, and is promoting cell proliferation; it is desirable to decrease the expression of HRM. In cancers where HRM is an inhibitor or suppressor and is controlling or decreasing cell proliferation, it is desirable to provide the protein or to increase the expression of HRM.

In one embodiment, where HRM is an inhibitor, HRM or a portion or derivative thereof may be administered to a subject to treat a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, an agonist which is specific for HRM may be administered to a subject to treat a cancer including, but not limited to, those cancers listed above.

In another further embodiment, a vector capable of expressing HRM, or a portion or a derivative thereof, may be administered to a subject to treat a cancer including, but not limited to, those cancers listed above.

In a further embodiment where HRM is promoting cell proliferation, antagonists which decrease the expression or activity of HRM may be administered to a subject to treat a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind HRM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HRM.

In another embodiment, a vector expressing the complement of the polynucleotide encoding HRM may be administered to a subject to treat a cancer including, but not limited to, those cancers listed above.

In yet another embodiment where HRM is promoting leukocyte activity or proliferation, antagonists which decrease the activity of HRM may be administered to a subject to treat an immune response. Such responses may be associated with AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In one aspect, antibodies which specifically bind HRM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HRM.

In another embodiment, a vector expressing the complement of the polynucleotide encoding HRM may be administered to a subject to treat an immune response including, but not limited to, those listed above

In one further embodiment, HRM or a portion or derivative thereof may be added to cells to stimulate cell proliferation. In particular, HRM may be added to a cell in culture or cells in vivo using delivery mechanisms such as liposomes, viral based vectors, or electroinjection for the purpose of promoting cell proliferation and tissue or organ regeneration. Specifically, HRM may be added to a cell, cell line, tissue or organ culture in vitro or ex vivo to stimulate cell proliferation for use in heterologous or autologous transplantation. In some cases, the cell will have been preselected for its ability to fight an infection or a cancer or to correct a genetic defect in a disease such as sickle cell anemia, β thalassemia, cystic fibrosis, or Huntington's chorea.

In another embodiment, an agonist which is specific for HRM may be administered to a cell to stimulate cell proliferation, as described above.

In another embodiment, a vector capable of expressing HRM, or a portion or a derivative thereof, may be administered to a cell to stimulate cell proliferation, as described above.

In other embodiments, any of the therapeutic proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other therapeutic agents. Selection of the agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment of the various disorders described above. Using this approach,

one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of HRM may be produced using methods which are generally known in the art. In particular, purified HRM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HRM.

Antibodies to HRM may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with HRM or any portion or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or portions used to induce antibodies to HRM have an amino acid sequence consisting of at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HRM amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to HRM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler et al. (1975) Nature 256:495-497; Kozbor et al. (1985) J Immunol Methods 81:31-42; Cote et al. (1983) Proc Natl Acad Sci 80:2026-2030; Cole et al. (1984) Mol Cell Biol 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with antigen specificity and biological activity can be used (Morrison et al. (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger et al. (1984) Nature 312:604-608; and Takeda et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce

HRM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton (1991) Proc Natl Acad Sci 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi et al. (1989) Proc Natl Acad Sci 86:3833-3837; Winter et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for HRM may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HRM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HRM epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding HRM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HRM may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HRM. Thus, complementary molecules or fragments may be used to modulate HRM activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding HRM.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of polynucleotides to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding HRM. These techniques are described both in Sambrook (supra) and in Ausubel (supra).

Genes encoding HRM can be turned off by transforming a cell or tissue with expression vectors

which express high levels of a polynucleotide or fragment thereof which encodes HRM. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions of the gene encoding HRM (signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee et al. (1994) In: Huber and Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HRM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HRM. Such DNA sequences may be incorporated into a wide variety of vectors with RNA polymerase promoters such as

T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino polymers (Goldman et al. (1997) Nature Biotechnol 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HRM, antibodies to HRM, mimetics, agonists, antagonists, or inhibitors of HRM. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for

formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Mack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding auxiliaries, if desired, to obtain tablets or dragee cores. Excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution. Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as oily injection suspensions. Lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly

concentrated solutions.

For topical or nasal administration, penetrants to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in a container and labeled for treatment of an indicated condition. For administration of HRM, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HRM or portions thereof, antibodies of HRM, agonists, antagonists or inhibitors of HRM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50.

Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form

employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide levels of the active moiety that produce or maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or proteins will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HRM may be used for the diagnosis of conditions or diseases characterized by expression of HRM, or in assays to monitor patients being treated with HRM, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for HRM include methods which utilize the antibody and a label to detect HRM in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring HRM are known in the art and provide a basis for diagnosing altered or abnormal levels of HRM expression. Normal or standard values for HRM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HRM under conditions for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of HRM expressed in subject samples are compared with the standard values from control and diseased samples. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HRM may be used for

diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HRM may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of HRM, and to monitor regulation of HRM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding HRM or closely related molecules, may be used to identify nucleic acid sequences which encode HRM. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding HRM, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the HRM encoding sequences. The hybridization probes of the invention may be DNA or RNA and derived from the polynucleotide of SEQ ID NOs:50-98 or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring HRM.

Means for producing specific hybridization probes for polynucleotides encoding HRM include the cloning of nucleic acid sequences encoding HRM or HRM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the RNA polymerases and the labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding HRM may be used for the diagnosis of conditions, disorders, or diseases which are associated with either increased or decreased expression of HRM. Examples of such conditions or diseases include adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers of the adrenal gland, bladder, bone, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, bone marrow, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and immune disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus,

multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and thyroiditis. The polynucleotides encoding HRM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, or other multiformat assays including microarrays to analyze fluids or tissues from patient biopsies to detect altered HRM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the polynucleotides encoding HRM may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The polynucleotides encoding HRM may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed; the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly different from that of a comparable control sample, the polynucleotides have hybridized with nucleic acids in the sample, and the presence of differentially expressed polynucleotides encoding HRM in the sample indicates the presence of the disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of HRM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes HRM, under conditions for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of an isolated polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ aggressive treatment earlier thereby preventing further

progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HRM may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two polynucleotides, one with sense orientation (5'->3') and another with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HRM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby et al. (1993) J Immunol Methods, 159:235-244; Duplaa et al. (1993) Anal Biochem 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an multiwell format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides may be used as targets on a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information may be used to determine gene function, understanding the genetic basis of disease, diagnosing disease, and in developing and in monitoring the activities of therapeutic agents.

In one embodiment, the microarray is prepared and used according to the methods described in PCT application WO95/11995, Lockhart et al. (1996; Nature Biotechnol 14:1675-1680) and Schena et al. (1996; Proc Natl Acad Sci 93:10614-10619), all of which are incorporated herein in their entirety by reference.

The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray, it may be preferable to use oligonucleotides which are only 7-10 nucleotides in length. The microarray may contain oligonucleotides which cover the known 5', or 3', sequence, or contain sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell or tissue type or to a normal, developmental, or

disease state. In certain situations it may be to use pairs of oligonucleotides on a microarray. The "pairs" will be identical, except for one nucleotide which is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from 2 to one million.

In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' or more preferably at the 3' end of the polynucleotide. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range for hybridization, and lack predicted secondary structure that may interfere with hybridization. In one aspect, the oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide, or any other solid support.

In another aspect, the oligonucleotides may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a gridded array analogous to a dot or slot blot apparatus may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. In yet another aspect, an array may be produced by hand or using available devices, materials, and machines (including multichannel pipetters or robotic instruments) and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other multiple from 2 to one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarrays, polynucleotides are extracted from a biological sample. The biological samples may be obtained from any bodily fluid (blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. To produce probes, the polynucleotides extracted from the sample are used to produce nucleic acid sequences which are complementary to the nucleic acids on the microarray. If the microarray consists of cDNAs, antisense RNAs (aRNA) are probes. Therefore, in one aspect, mRNA is used to produce cDNA which, in turn and in the presence of fluorescent nucleotides, is used to produce fragment or oligonucleotide aRNA probes. These fluorescently labeled probes are incubated with the microarray so that the probe sequences hybridize to the cDNA oligonucleotides of the microarray. In another aspect, complementary nucleic acid sequences are used as probes and can also include polynucleotides, fragments, complementary, or antisense sequences produced using restriction enzymes, PCR technologies, and oligolabeling kits which are well known in the art.

Incubation conditions are adjusted so that hybridization occurs with precise complementary matches

or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies or functional analysis of the sequences, mutations, variants, or polymorphisms among samples (Heller et al. (1997) Proc Natl Acad Sci 94:2150-55).

In another embodiment of the invention, the nucleic acid sequences which encode HRM may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries as reviewed in Price (1993; Blood Rev 7:127-134) and Trask (1991; Trends Genet 7:149-154).

Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at Online Mendelian Inheritance in Man (OMIM). Correlation between the location of the gene encoding HRM on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The polynucleotides of the invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The polynucleotide of the invention may also be used to detect differences in the chromosomal location due to

translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, HRM, its catalytic or immunogenic portions or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The portion employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between HRM and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HRM, or portions thereof, and washed. Bound HRM is then detected by methods well known in the art. Purified HRM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HRM specifically compete with a test compound for binding HRM. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HRM.

In additional embodiments, the polynucleotides which encode HRM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of polynucleotides that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the invention and are not included for the purpose of limiting the invention.

EXAMPLES

For purposes of example, the preparation and sequencing of the LNODNOT03 cDNA library, from which Incyte Clones 1572888, 1573677, 1574624, and 1577239 were isolated, is described. Preparation and sequencing of cDNAs in libraries in the LIFESEQ database (Incyte Genomics, Palo Alto CA) have varied over time, and the gradual changes involved use of kits, plasmids, and machinery available at the particular time the library was made and analyzed.

I LNODNOT03 cDNA Library Construction

The LNODNOT03 cDNA library was constructed using 1 µg of polyA RNA isolated from lymph node tissue removed from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy.

Microscopic examination showed that the tissue was extensively necrotic with 10% viable tumor. The invasive grade 3/4 squamous cell carcinoma had formed a mass in the right lower lobe of the lung which had invaded into, but not through, the visceral pleura. Focally, the tumor had obliterated the bronchial lumen although the bronchial margin was negative for dysplasia/neoplasm. One of two intrapulmonary, one of four inferior mediastinal (subcarinal), and two of eight superior mediastinal lymph nodes were metastatically involved. Patient history included hemangioma and tobacco use; the patient was taking Doxycycline, a tetracycline, to treat an infection.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NY). The lysate was centrifuged over a 5.7 M CsCl cushion using an SW28 rotor in a L8-70M ultracentrifuge (Beckman Coulter, Fullerton CA) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol, pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and treated with DNAse at 37C. Extraction and precipitation were repeated as before. The mRNA was isolated using the OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERSCRIPT plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (APB), and those cDNAs exceeding 400 bp were ligated into pINCY plasmid (Incyte Genomics). The plasmid was subsequently transformed into DH5α competent cells (Life Technologies).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Qiagen). This kit enabled the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile TERRIFIC BROTH (BD Biosciences, Sparks MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after incubation for 19 hours, the cells were lysed with 0.3 ml of lysis buffer and precipitated using isopropanol, and 3) the plasmid pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

The cDNAs were prepared using a MICROLAB system (Hamilton) in combination with DNA ENGINE thermal cyclers (MJ Research), sequenced by the method of Sanger and Coulson (1975, J Mol Biol 94:441f) using 377 PRISM DNA sequencing systems (ABI); and reading frame was determined.

III Homology Searching of cDNA Clones and Their Deduced Proteins

The polynucleotides and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain

previously identified and annotated sequences, were searched for regions of homology using BLAST, which stands for Basic Local Alignment Search Tool (Altschul (1993) J Mol Evol 36:290-300; Altschul et al. (1990) J Mol Biol 215:403-410).

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms such as the one described in Smith et al. (1992, Protein Engineering 5:35-51) could have been used when dealing with primary sequence patterns and secondary structure gap penalties. The sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-14} for peptides.

Incyte polynucleotides were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam); and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp) for homology.

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled polynucleotide to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook, supra).

Analogous computer techniques use BLAST to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ database (Incyte Genomics). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding HRM occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V Extension of HRM Encoding Polynucleotides

The nucleic acid sequence of an Incyte Clone disclosed in the Sequence Listing was used to design oligonucleotide primers for extending a partial sequence to full length. One primer was synthesized to initiate extension in the antisense direction, and the other was synthesized to extend sequence in the sense direction. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO software (Molecular Insights), or another program, to be about 22 to about 30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures of about 68 to about 72C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (Life Technologies) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (ABI) and thoroughly mixing the enzyme and reaction mix. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the DNA ENGINE thermal cycler (MJ Research) and the following parameters: Step 1, 94C for 1 min (initial denaturation); Step 2, 65C for 1 min; Step 3, 68C for 6 min; Step 4, 94C for 15 sec; Step 5, 65C for 1 min; Step 6, 68C for 7 min; Step 7, repeat step 4-6 for 15 additional cycles; Step 8, 94C for 15 sec; Step 9, 65C for 1 min; Step 10, 68C for 7:15 min; Step 11, repeat step 8-10 for 12 cycles; Step 12, 72C for 8 min; and Step 13, hold at 4C.

A 5-10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK kit (Qiagen), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16C. Competent E. coli cells (in 40 μl of media) were transformed

with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook, supra). After incubation for one hour at 37C, the E. coli mixture was plated on Luria Bertani (LB)-agar (Sambrook, supra) containing 2x carbenicillin (Carb). The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of a commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions: Step 1, 94C for 60 sec; Step 2, 94C for 20 sec; Step 3, 55C for 30 sec; Step 4, 72C for 90 sec; Step 5, repeat steps 2-4 for an additional 29 cycles; Step 6, 72C for 180 sec; and Step 7, hold at 4C.

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and clones were selected, ligated into plasmid, and sequenced.

In like manner, a genomic library and a polynucleotide selected from SEQ ID NOs:50-98 is used to obtain 5' regulatory sequences using the procedure above.

VI Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NOs:50-98 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO software (Molecular Insights), labeled by combining 50 pmol of each oligomer and 250 μCi of [γ-^{32}P] adenosine triphosphate (APB) and T4 polynucleotide kinase (NEN Life Science Products, Acton MA). The labeled oligonucleotides are purified using SEPHADEX G-25 superfine resin column (APB). A aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; NEN Life Science Products).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to NYTRANPLUS membranes (Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR film (Eastman Kodak, Rochester NY) is exposed to the blots in a PHOSPHOIMAGER cassette (APB) for several hours, hybridization patterns are compared visually.

VII Microarrays

To produce oligonucleotides for a microarray, SEQ ID NOs:50-98 are examined using a computer algorithm which starts at the 3' end of the polynucleotide. The algorithm identified oligomers of defined length that are unique to the gene, have a GC content within a range for hybridization, and lack predicted secondary structure that would interfere with hybridization. The algorithm identifies approximately 20 sequence-specific oligonucleotides of 20 nucleotides in length (20-mers). A matched set of oligonucleotides are created in which one nucleotide in the center of each sequence is altered. This process is repeated for each gene in the microarray, and double sets of twenty 20 mers are synthesized and arranged on the surface of the silicon chip using a light-directed chemical process (described in PCT/WO95/11995).

In the alternative, a chemical coupling procedure and an ink jet device are used to synthesize oligomers on the surface of a substrate (PCT/WO95/251116). In another alternative, a gridded array is used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical, or chemical bonding procedures. A typical array may be produced by hand or using available materials and machines and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots. After hybridization, the microarray is washed to remove nonhybridized probes, and a scanner is used to determine the levels and patterns of fluorescence. The scanned image is examined to determine degree of complementarity and the relative abundance/expression level of each sequence in the microarray.

VIII Complementary Polynucleotides

Sequence complementary to the sequence encoding HRM, or any part thereof, is used to detect, decrease or inhibit expression of naturally occurring HRM. Although use of oligonucleotides comprising from about 15 to about 30 base-pairs is described, the same procedure is used with smaller or larger sequence fragments. Oligonucleotides are designed using OLIGO software (Molecular Insights) and the coding sequence of SEQ ID NOs:50-98. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the transcript encoding HRM.

IX Expression of HRM

Expression of HRM is accomplished by subcloning the cDNAs into vectors and transforming the vectors into host cells. In this case, the cloning vector is also used to express HRM in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique

restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of HRM into the bacterial growth media which can be used directly in the following assay for activity.

X Demonstration of HRM Activity

HRM can be expressed in a mammalian cell line such as DLD-1 or HCT116 (ATCC) by transforming the cells with a eukaryotic expression vector encoding HRM. Eukaryotic expression vectors are commercially available and the techniques to introduce them into cells are well known to those skilled in the art. The effect of HRM on cell morphology may be visualized by microscopy; the effect on cell growth may be determined by measuring cell doubling-time; and the effect on tumorigenicity may be assessed by the ability of transformed cells to grow in a soft agar growth assay (Groden (1995) Cancer Res 55:1531-1539).

XI Production of HRM Specific Antibodies

HRM that is purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. In the alternative, an amino acid sequence deduced from SEQ ID NOs:50-98 is analyzed using LASERGENE software (DNASTAR, Madison WI) to determine regions of high immunogenicity, and an oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of epitope, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel (supra).

Typically, the oligopeptides are 15 residues in length, synthesized using a 431A Peptide synthesizer (ABI) using Fmoc-chemistry, and coupled to keyhole limpet hemocyanin (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (Ausubel supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the protein to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio iodinated, goat anti-rabbit IgG.

XII Purification of Naturally Occurring HRM Using Specific Antibodies

Naturally occurring or recombinant HRM is substantially purified by immunoaffinity chromatography using antibodies specific for HRM. An immunoaffinity column is constructed by covalently coupling HRM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE resin (APB). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HRM is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HRM (e.g., high ionic strength buffers in the presence of

detergent). The column is eluted under conditions that disrupt antibody/protein binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HRM is collected.

XIII Identification of Molecules Which Interact with HRM

HRM, or biologically active portions thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al. (1973) Biochem J 133:529-39). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HRM, washed and any wells with labeled HRM complex are assayed. Data obtained using different concentrations of HRM are used to calculate values for the number, affinity, and association of HRM with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.